51

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/746,583	12/22/2000	William H. Rockenbeck	M61.12-0320	1587

7590 10/06/2004

Theodore M. Magee
WESTMAN, CHAMPLIN & KELLY, P.A.
International Center-Suite 1600
900 Second Avenue South
Minneapolis, MN 55402-3319

EXAMINER	
STORM, DONALD L	
ART UNIT	**PAPER NUMBER**
2654	

DATE MAILED: 10/06/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/746,583	ROCKENBECK ET AL.	
	Examiner	Art Unit	
	Donald L. Storm	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *22 December 2000*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-31* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-3,12,13,15 and 17-31* is/are rejected.

7)☒ Claim(s) *4-11,14 and 16* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *22 December 2000* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date 2.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Claim Informalities

1. Claims 4-11, 14, 16, are objected to as being (directly or indirectly) dependent upon a rejected base claim. See MPEP § 608.01(n)V. The claim(s) would be allowable over the prior art of record if rewritten to include all of the limitations of the base claim and any intervening claims. The claims should also be rewritten to overcome any objections or rejections under 35 U.S.C. 112, especially as appearing in this Office action. Certain assumptions that make the limitations clear have been considered for the claims, as described next or elsewhere in this Office action.

2. The preamble of claim 1 is objected to under 37 CFR 1.75(a) because the invention established by the preamble is not carried out by the limitations in the body of the claim. The preamble establishes a claim to a method of storing speech information; however, the subject matter described by the limitations in the body of the claim is directed solely toward steps for accumulating a particular kind of sum. There are no steps that store anything. Thus, the body of the claim is unconnected to the storing method set forth in the preamble, but the body of the claim is able to stand alone. The disconnect leaves an artisan uncertain (1) whether the received speech signal must be stored after it has been received or perhaps it is now being received from some storage, (2) whether the claim somehow further limits the speech signal features, frames, etc. by the functionality of storing, (3) or whether the claimed invention includes only speech signal information that is stored sometime in the uncertain future according to some unspecified storage method or means. It is confusing to establish a certain objective to be achieved by a method, but to define the method only by steps that do not accomplish that objective.

3. Claim 1, and by dependency claims 2-14, are objected to under 37 CFR 1.75(a) because the meaning of the word "each" (line 6) needs clarification. It may be unclear as to what element this word "each" refers. In the grammatical construction "each of a set of frames," the word "each" would seem to an ellipsis of "each frame of a set of frames"; however, in the grammatical construction "one feature value for each," the word each would seem to be ellipsis of "one feature value for each feature value." To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase to refer to --each frame of a set of frames--.

4. Claim 8, and by dependency claim 9, are objected to under 37 CFR 1.75(a) because the meaning of the phrase "the training operations" (line 3) needs clarification. Because no training operations were previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase to refer to --training operations--.

5. Claim 12, and by dependency claim 13, are objected to under 37 CFR 1.75(a) because the meaning of the phrase "the states that form the alignment unit" (line 4) needs clarification. If the states that form the alignment unit are a plurality of the state of an alignment unit as previously recited in claim 1, the same phrasing should be used. If not, then further clarification is needed. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase to refer to --states that form the alignment unit--.

6. The preamble of claim 15, and by dependency claims 16-19, are objected to under 37 CFR 1.75(a) because the invention established by the preamble is not carried out by the limitations in

the body of the claim. The preamble establishes a claim to a speech recognition system and to recognizing linguistic units; however, the subject matter described by the limitations in the body of the claim is only directed toward training acoustic models and includes identifying alignment units. Thus, the body of the claim is unconnected to the speech recognition and to the linguistic units set forth in preamble, but the body of the claim is able to stand alone. The disconnect leaves an artisan uncertain (1) whether the alignment units are linguistic alignment units, (2) whether the scope of the invention includes representing all claimed acoustic models whether or not the models are ever used for recognition and whether or not the acoustic models correspond to speech, or (3) whether the claimed invention includes only acoustic models that are speech models and that are used sometime in the uncertain future for recognition according to some unspecified recognition method or means. It is confusing to establish a certain objective to be achieved by a system, but to define the systems only by means that do not accomplish that objective.

7. Claim 19 is objected to under 37 CFR 1.75(a) because the meaning of the phrase "the states of the word" (last line) needs clarification. Because no states were previously associated with words, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase to refer to --states of the word--.

8. The preamble of claim 20, and by dependency claims 21-25, are objected to under 37 CFR 1.75(a) because the invention established by the preamble is not carried out by the limitations in the body of the claim. The preamble establishes a claim to a method of aligning frames of a speech signal; however, the subject matter described by the limitations in the body of the claim is only directed toward frames associated with an alignment unit. Thus, the body of the claim is

unconnected to the frames of speech set forth in preamble, but the body of the claim is able to stand alone. The disconnect leaves an artisan uncertain (1) whether the alignment units associated with frames are speech alignment units, (2) whether the scope of the invention includes representing all claimed frames associated with alignment units, whether or not the frames correspond to speech, or (3) whether the claimed invention includes only frames that might be speech sometime in the uncertain future according to some unspecified framing method or means. It is confusing to establish a certain objective to be achieved by a method, but to define the method only by steps that do not accomplish that objective.

9. Claim 20, and by dependency claims 21-25, are objected to under 37 CFR 1.75(a) because the meaning of the phrase "by the decoder" needs clarification. Because no decoder was previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has not assigned patentable weight to the phase --by the decoder--.

Claim Rejections - 35 USC § 101

10. The following is a quotation of 35 U.S.C. 101:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

11. Claim 20 is rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

12. Regarding claim 20, the body of the claim recites procedures that describe steps of a mathematical algorithm. A purely mathematical algorithm is nonstatutory despite the fact that it might inherently have some usefulness. Taken as a whole, the process of the claim is drawn to a mathematical method that begins with a set of frames already existing and manipulates that as data to produce other data as aligned states and frames, however, with no application of the resulting aligned data or of any intermediate data. For such subject matter to be statutory, the claimed process must actively and positively recite a practical application of the algorithm. A mathematical algorithm that simply manipulates data is nonstatutory; however, a claimed significant use of the results could be statutory.

The preamble describes that the frames are "of a speech signal" as a practical application; however, it is not clear how the invention is intended for this application. The recited "frames of a speech signal" in the claim's preamble merely specifies an intended use for the invention, and appears to direct the invention to data that is already prepared. For such subject matter to be statutory, the claimed process must actively and positively recite a practical application of the algorithm, such as representation of speech as in claim 21. Outputting a model provides a useful, concrete, and tangible result, namely a representation of the mathematics that is momentarily fixed and relied upon for outputting. An alternate application that may be statutory would be adding the positive recitation of receiving a speech signal, as in claim 1.

Taken as a whole, the claim is drawn to a mathematical method for manipulating data to produce other data with no significant application of the result. The algorithm manipulates symbols. A mathematical algorithm that simply manipulates data or symbols is nonstatutory. All claim limitations have been considered, and the claimed methods have been found nonstatutory as

a mathematical algorithm produces a set of numbers in one format from another set of numbers in another format, and without claiming a practical application.

Claim Rejections - 35 USC § 103

13. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

Takagi '094 and Takagi '223

14. Claims 1-3, 12-13, 15, and 17-31 are rejected under 35 U.S.C. 103(a) as being unpatentable over Takagi et al. [US Patent 5,651,094, Takagi '094] in view of Takagi [US Patent 5,819,223, Takagi '223].

15. Regarding claim 26, Takagi '094 [at column 8, lines 21-26] makes obvious a frame to state alignment embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

a decoder that identifies a sequence of alignment units from a speech signal [at column 8, lines 58-66, as the conventional analyzer that converts input speech into a time sequence of feature vectors];

the decoder associates sets of frames from the speech signal with the alignment units [at column 2, lines 14-36, as the analyzer extracts features vectors registered as reference patterns as representing speech of a standard speaker using discrete times as a frame];

an aligner that aligns an alignment unit with frames in the associated set of frames [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors input speech and features vectors representing time frames of the reference pattern].

Although Takagi '094 [at column 2, lines 54-55] suggests an HMM matching method, Takagi '094 does not provide details of HMM matching. In particular, Takagi '094 does not explicitly describe aligning acoustic states with frames.

On the other hand, Takagi '223 [for the sixth device] makes obvious a frame to state alignment embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

a decoder that identifies a sequence of alignment units from a speech signal [at column 4, lines 31-33, as the analysis unit that converts input speech into a time sequence of feature vectors];

the decoder associates sets from the speech signal with the alignment units [at column 4, lines 31-42, as the analysis unit converts input speech into feature vectors X(t) represented at a discrete time instant];

a trainer controller that identifies acoustic states for the alignment units [at column 9, lines 56-63, as the acoustic unit allows reception by separating the sequence of a state of HMM into the acoustic unit];

an aligner that aligns the acoustic states of an alignment unit [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns].

Takagi '223 [at column 10, lines 26-26] points out several advantages to be gained by using the popular HMM representation of speech, and explicitly describes details of matching using the states of an HMM model. In view of the similarities of concept and operations between Takagi '094 and Takagi '223, incorporating the concepts of one into the other would have been obvious to one of ordinary skill in the art of speech processing at the time of invention. In particular, to the extent that Takagi '094 does not necessarily include states in HMM reference patterns used for matching, it would have been obvious to an artisan to use the concepts described by Takagi '223 at least by alignment of Takagi '094's frames with states of the HMM, as Takagi '223 describes, because the HMM structure models any and all contents of speech, and it has no nonlinear extension function in time.

16. Regarding claim 27, Takagi '094 also describes:

an acoustic model that is used b the decoder to identify the sequence of alignment units form the speech signal [at column 2, line 14-column 3, line 9, as features vectors registered as reference patterns as representing speech of a standard speaker are used for matching feature vectors of respective input speech converted by the analyzer to output the reference pattern which gives a minimum distance from the input speech].

17. Regarding claim 28, Takagi '094 also describes:

the frame alignment system forms part of a model adaptation system for adapting the acoustic model [see Fig. 5, items 1, 2, 22, 55, 2' and their descriptions especially at column 3, lines 10-11, as carrying out the matching for adaptation or learning].

18. Regarding claim 29, Takagi '094 also describes:

a feature extractor that generates a feature vector for each frame of the speech signal [at column 2, lines 14-34, as the spectral analysis processes that convert input speech into feature vectors using discrete times as a frame];

each feature vector comprising a plurality of dimension values for respective dimensions of the feature vector [at column 2, lines 20-27, as the method obtains multidimensional vectors with various parameters including a spectrum, etc.].

19. Regarding claim 30, Takagi '094 also describes:

a dimension sum storage for storing sums [at column 3, line 67-column 4, line 26, as N(c) and S(c) after adding X(i,c) and the mean values in the respective acoustic categories];

the sums are dimension sums for each state [at column 2, lines 59-64, as the feature vectors X(i,c) represent vector component c at time frame i];

each dimension sum is associated with a dimension of the feature vectors [at column 2, lines 59-64, as the feature vectors X(i,c) represent vector component c];

the sum is formed by adding the dimension values that are found in the feature vectors [at column 2, lines 59-64, as the feature vectors X(i,c) represent vector component c];

the feature vectors are associated with frames *{exmr: were generated for frames}* [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors input speech and features vectors representing time frames of the reference pattern];

and Takagi '223 describes:

the alignment was with the state [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns].

20. Regarding claim 31, Takagi '094 also describes:

a model adapter that uses the dimension sums to adapt the acoustic model [at column 12, line 66-column 13, line 10, as mean values I(p,c) and M(p,c) determines an adaptation vector and adapts the reference patterns].

21. Regarding claim 20, Takagi '094 [at column 8, lines 21-26] makes obvious a frame to state alignment embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

identifying alignment units [at column 8, lines 58-66, as converting input speech into feature vectors];

the alignment units correspond to a sequence of linguistic units [at column 2, lines 37-38, as feature vectors are registered in units of words];

identifying a set of frames that are associated with each linguistic unit [at column 2, lines 14-36, as using discrete times as a frame and extracting feature vectors registered in units of words as reference patterns representing speech of a standard speaker];

for each of the alignment units, aligning the frames associated with an alignment unit with [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors input speech and features vectors representing time frames of the reference pattern].

Although Takagi '094 [at column 2, lines 54-55] suggests an HMM matching method, Takagi '094 does not provide details of HMM matching. In particular, Takagi '094 does not explicitly describe aligning acoustic states with frames.

On the other hand, Takagi '223 [for the sixth device] makes obvious a frame to state alignment embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

identifying alignment units [at column 4, lines 31-33, as convert input speech into feature vectors];

the alignment units correspond to a sequence of linguistic units [at column 1, lines 45-50, as a time series of feature vectors were memorized as a plurality of word reference patterns];

for each of the alignment units, identifying the states associated with the alignment unit [at column 6, lines 26-32, as typical sequences of plural states of an HMM are separated individually so that the reference patterns and all contents of an utterance can be received];

for each of the alignment units, aligning the acoustic states associated with an alignment unit [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns].

Takagi '223 [at column 10, lines 26-26] points out several advantages to be gained by using the popular HMM representation of speech, and explicitly describes details of matching using the states of an HMM model. In view of the similarities of concept and operations between Takagi '094 and Takagi '223, incorporating the concepts of one into the other would have been

obvious to one of ordinary skill in the art of speech processing at the time of invention. In particular, to the extent that Takagi '094 does not necessarily include states in HMM reference patterns used for matching, it would have been obvious to an artisan to use the concepts described by Takagi '223 at least by alignment of Takagi '094's frames with states of the HMM, as Takagi '223 describes, because the HMM structure models any and all contents of speech, and it has no nonlinear extension function in time.

22. Regarding claim 21, Takagi '223 also describes:

the method is part of a process of associating feature vectors that represent the speech signal with states of words [at column 9, lines 10-22, as the results can be applied to word spotting by using, for example, a plurality of states of the HMM].

23. Claim 22 sets forth additional limitations similar to limitations set forth in claim 29. Takagi '094 and Takagi '223 describe and make obvious the additional limitations as indicated there.

24. Claim 23 sets forth additional limitations similar to limitations set forth in claims 30 and 31. Takagi '094 and Takagi '223 describe and make obvious the additional limitations as indicated there.

25. Claim 24 sets forth additional limitations comprising the functionality associated with using the system recited in claim 31. Takagi '094 and Takagi '223 describe and make obvious these additional limitations as indicated there.

26. Regarding claim 25, Takagi '094 also describes:

using the dimension sums to change the parameters of a initial acoustic model to form an adapted acoustic model [at column 13, lines 1-47, as add an adaptation vector based on M(p,c) to adapt the reference patterns thereby generating new reference patterns].

27. Regarding claim 15, Takagi '094 [at column 1] makes obvious a system generally applied to speech recognition as an embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

an acoustic model [at column 2, lines 14-40, as stored reference patterns of speech of a standard speaker converted into a sequence of speech features];

a decoder to identify alignment units in a speech signal [at column 8, lines 58-66, as the conventional analyzer that converts input speech into a time sequence of feature vectors];

the acoustic model is used to identify them [at column 2, lines 14-36, as reference patterns are registered as representing speech of a standard speaker using discrete times as the analyzer extracts features vectors];

an aligner that aligns the identified alignment unit with frames of the speech signal [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors of input speech and features vectors representing time frames of the reference pattern];

a dimension sum storage that stores sums [at column 3, line 67-column 4, line 26, as N(c) and S(c) after adding X(i,c) and the mean values in the respective acoustic categories];

the sums are feature dimension sums that are associated with alignment units [at column 2, lines 59-64, as the feature vectors X(i,c) represent vector component c at time frame i];

each sum updated by summing dimension values from feature vectors [at column 2, lines 59-64, as the feature vectors X(i,c) represent vector component c];

the feature vectors are assigned to the aligned frames [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors input speech and features vectors representing time frames of the reference pattern];

a sum updated before a sufficient number of frames are available [at column 4, lines 10-36, as S(c) = S(c) + X(i,c) repeatedly until i and j are decremented to 0 then, if both are zero calculate the mean value of the completed repeated steps by dividing S(c)];

the sum is either an insufficient number of frames or a sufficient number of frames to train the acoustic model [at column 12, line 66-column 13, line 10, as use the mean values to calculate an adaptation vector to adapt the reference patterns];

a model adapter that uses the feature dimension sums to train the acoustic model [at column 12, line 66-column 13, line 10, as mean values I(p,c) and M(p,c) determines an adaptation vector and adapts the reference patterns].

Although <u>Takagi '094</u> [at column 2, lines 54-55] suggests an HMM matching method, <u>Takagi '094</u> does not provide details of HMM matching. In particular, <u>Takagi '094</u> does not explicitly describe aligning acoustic states with frames.

On the other hand, <u>Takagi '223</u> [for the sixth device] makes obvious a frame to state alignment embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

an acoustic model [at column 4, lines 59-62, as the acoustic unit registered as an HMM for reference speaker speech];

a decoder that identifies alignment units in the speech signal [at column 4, lines 31-33, as the analysis unit that converts input speech into a time sequence of feature vectors];

the identification uses the acoustic model [at column 9, lines 56-63, as the acoustic unit allows reception by separating the sequence of a state of HMM into the acoustic unit];

an aligner that aligns the acoustic states of an alignment unit [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns];

a model adapter to train the acoustic model [at column 2, line 67-column 3, line 4, as an adaptation unit for making correction of the feature vectors of the reference pattern by using the mean vectors].

Takagi '223 [at column 10, lines 26-26] points out several advantages to be gained by using the popular HMM representation of speech, and explicitly describes details of matching using the states of an HMM model. In view of the similarities of concept and operations between Takagi '094 and Takagi '223, incorporating the concepts of one into the other would have been obvious to one of ordinary skill in the art of speech processing at the time of invention. In particular, to the extent that Takagi '094 does not necessarily include states in HMM reference patterns used for matching, it would have been obvious to an artisan to use the concepts described by Takagi '223 at least by alignment of Takagi '094's frames with states of the HMM, as Takagi '223 describes, because the HMM structure models any and all contents of speech, and it has no nonlinear extension function in time.

28. Regarding claim 17, Takagi '094 also describes:

an initial acoustic model [at column 2, lines 14-36, as reference patterns converted from speech of a standard speaker into a sequence of speech features];

train the acoustic model by adapting the parameters of a initial acoustic model to form a new version of the acoustic model [at column 13, lines 1-47, as add an adaptation vector based on M(p,c) to adapt the reference patterns thereby generating new reference patterns].

29. Regarding claim 19, Takagi '094 also describes:

the decoder assigns frames of speech signals to words [at column 2, lines 14-36, as using discrete times as a frame and extracting feature vectors registered in units of words];

and Takagi '223 describes:

the aligner aligns with states of words [at column 9, lines 10-22, as the results can be applied to word spotting by using, for example, a plurality of states of the HMM].

30. Regarding claim 1, Takagi '094 [at abstract] makes obvious a method of retraining a speech model recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

receiving a speech signal [see Fig. 5, item INPUT, 1, and their descriptions especially at column 2, line 14, as input speech];

decoding it to identify a sequence of alignment units [at column 8, lines 58-66, as the conventional analyzer that converts input speech into a time sequence of feature vectors];

decoding it based on a speech model [at column 2, lines 14-36, as features vectors extracted by the analyzer are registered as reference patterns as representing speech of a standard speaker];

identifying a feature value for each of a set of frames of a speech signal [at column 2, lines 14-34, as the spectral analysis processes that convert input speech into feature vectors using discrete times as a frame];

aligning an alignment unit from their sequence with a frame in the their set [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors input speech and features vectors representing time frames of the reference pattern].

Although Takagi '094 [at column 2, lines 54-55] suggests an HMM matching method, Takagi '094 does not provide details of HMM matching. In particular, Takagi '094 does not explicitly describe aligning acoustic states with frames.

On the other hand, Takagi '223 [for the sixth device] makes obvious a frame to state alignment embodiment recognizable as a whole to one versed in the art by explicitly describing the content and functionality of the recited limitations as the following terminology:

decoding a speech signal to identify a sequence of alignment units from a speech signal [at column 4, lines 31-33, as the analysis unit that converts input speech into a time sequence of feature vectors];

the decoding is based on the speech model [at column 9, lines 56-63, as the acoustic unit allows reception by separating the sequence of a state of HMM into the acoustic unit];

aligning a state of an alignment unit [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns].

Takagi '223 [at column 10, lines 26-26] points out several advantages to be gained by using the popular HMM representation of speech, and explicitly describes details of matching using the states of an HMM model. In view of the similarities of concept and operations between Takagi '094 and Takagi '223, incorporating the concepts of one into the other would have been obvious to one of ordinary skill in the art of speech processing at the time of invention. In particular, to the extent that Takagi '094 does not necessarily include states in HMM reference patterns used for matching, it would have been obvious to an artisan to use the concepts described by Takagi '223 at least by alignment of Takagi '094's frames with states of the HMM, as Takagi '223 describes, because the HMM structure models any and all contents of speech, and it has no nonlinear extension function in time.

Takagi '094 also describes:

before receiving enough frames, adding in the identified feature value to a feature value sum [at column 4, lines 10-36, as S(c) = S(c) + X(i,c) repeatedly until i and j are decremented to 0 then, if both are zero calculate the mean value of the completed repeated steps by dividing S(c)];

the number received was not enough to begin retraining [at column 12, line 66-column 13, line 10, as use the mean values to calculate an adaptation vector to adapt the reference patterns];

and Takagi '223 describes

the sum is associated with the aligned state [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns].

31. Regarding claim 2, Takagi '094 also describes:

the speech signal comprises a signal utterance [at column 1, line 19, as an unknown utterance].

32. Regarding claim 3, Takagi '094 also describes:

repeating the steps of identifying, decoding, aligning, and adding for each of a plurality of utterances [at column 1, lines 43-column 4, line 26, as the conventional technique of "adaptation," converting, storing, matching, summing, and dividing employing a few utterances].

33. Regarding claim 12, Takagi '094 also describes:

assigning frames to alignment units [at column 2, lines 43-64, as matching the sequence of feature vectors of input speech and features vectors representing time frames of the reference pattern];

aligning the alignment unit with frames assigned to the alignment unit [at column 2, lines 43-64, as a matching units that matches the sequence of feature vectors input speech and features vectors representing time frames of the reference pattern];

and Takagi '223 describes:

aligning states that form an alignment unit [at column 4, line 60-column 5, line 10, as a preliminary matching unit makes alignment between the series X(t) and the two state HMM of the reference patterns].

To the extent that Takagi '094 does not necessarily include states in HMM reference patterns used for matching, it would have been obvious to an artisan to use the concepts described

by Takagi '223 at least by alignment of Takagi '094's frames with states of the HMM, as Takagi '223 describes.

34. Regarding claim 13, Takagi '094 also describes:

the alignment unit is a word [at column 2, lines 14-36, as extracting feature vectors registered in units of words].

Takagi '094 and Takagi '223 and Gould

35. Claim 18 is rejected under 35 U.S.C. 103(a) as being unpatentable over Takagi et al. [US Patent 5,651,094, Takagi '094] in view of Takagi [US Patent 5,819,223, Takagi '223] and Gould et al. [US Patent 5,920,837].

36. Regarding claim 18, Takagi '094 and Takagi '223 describe and make obvious the included claim elements as indicated elsewhere in this Office action.

Throughout Takagi '094, the reference also describes algorithms for operating devices and apparatuses to accomplish the functions of a model adapter and a decoder. However, neither Takagi '094 nor Takagi '223 explicitly describes computer-executable instructions.

Gould [at columns 11-17] also describes a speech recognition system with model adaptation. Gould provides some details of a configuration for computer processors, as follows:

the decoder [at column 11, lines 28-35, as DSP operations of deriving the parameter vector in sound board circuitry];

the model adapter is a set of computer-executable instructions that are processed on a different thread from the decoder [at column 10, lines 20-54, as Adaptive Training instructions are loaded in RAM and executed by the CPU that is included in a computer].

To the extent that computer-executable code is not necessarily in Takagi '094's system or Takagi '223's system, the many teachings of Takagi '094 would have made it obvious to one of ordinary skill in the art of computer programming at the time of invention to install code configured in a CPU and sound board circuitry as described by Gould and automatically execute Takagi '094's and Takagi '223's algorithms on hardware because programmed processor implementation would eliminate tedious manual calculation of repetitive operations of the algorithms.

Conclusion

37. The following references here made of record are considered pertinent to applicant's disclosure:

Tzirkel-Hancock [US Patent 5,907,825] describes speech recognition implemented on a computer that allows a user to continuously build/update word models.

Lee, Chin-Hui, "Adaptive Compensation for Robust Speech Recognition," Proc. 1997 IEEE Workshop on Automatic Speech Recognition and Understanding, 1997., 14-17 Dec 1997, pp. 357-364, describes adaptive feature and model compensation which modifies either recognition feature vectors, recognition models, or both.

38. Any response to this action should be mailed to:

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

or faxed to:

(703) 872-9306, (for formal communications intended for entry)

Or:

(703) 872-9306, (for informal or draft communications, and please label "PROPOSED" or "DRAFT")

Patent Correspondence delivered by hand or delivery services, other than the USPS, should be addressed as follows and brought to U.S. Patent and Trademark Office, 220 20th Street S., Customer Window, Crystal Plaza Two, Lobby, Room 1B03, Arlington, VA, 22202

39. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Donald L. Storm, of Art Unit 2654, whose telephone number is (703) 305-3941. The examiner can normally be reached on weekdays between 8:00 AM and 4:30 PM Eastern Time. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (703) 305-9645.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 703-305-3028 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: ebc@uspto.gov. For general information about the PAIR system, see http://pair-direct.uspto.gov.

Donald L. Storm

Donald L. Storm
Patent Examiner
Art Unit 2654

September 30, 2004